Research Study

West Nile Virus Infection Risk is Associated with Combined Sewer Overflow Streams in Urban Atlanta, Georgia

Gonzalo M. Vazquez Prokopec [1]*, Jodi L. Vanden Eng[2], Rosmarie Kelly[3], Daniel G. Mead [4], Priti Kolhe [5], James Howgate [5], Uriel Kitron [1], Thomas R. Burkot[2]

[1] Emory University, Atlanta, Georgia, USA;[2] Centers for Disease Control and Prevention, Atlanta, Georgia, USA; [3] Georgia Division of Public Health, Atlanta; Georgia, USA [4] University of Georgia, Athens, Georgia, USA; [5] Fulton Department of Health and Wellness, Atlanta, Georgia, USA.

Supplemental Material

Supplemental Material and Methods

Tests of WNV infection

Sections of brain stem from dead birds and pooled mosquitoes (≤ 25 mosquitoes/pool) were homogenized in 1 ml virus isolation media (Minimum Essential Medium supplemented with 1,000 U penicillin G, 1 mg streptomycin, 0.25 mg gentamicin sulfate, 0.5 mg kanamycin monosulfate, 2.5 ug/ml amphotericin B, and 1% bovine serum albumin) using a Mini-BeadBeater 8 (BioSpec Products, Inc.) or a Qiagen7 Mixer Mill 300 (Valencia, CA) then clarified by centrifugation (10 min @ 7,200 x G). A 100 µl aliquot of the resulting supernatant fluid from each sample was inoculated onto a separate well of a 12-well plate with

confluent 2-day-old Vero Middle America Research Unit (MARU, [Vero M]) cell culture monolayer and incubated at 37 C in a 5% CO_2 atmosphere. Cell cultures were examined daily for 7 days for evidence of cytopathic effects (CPE). Samples not displaying CPE within the first 7 days were subpassaged onto 2-day-old Vero M cells and observed for an additional 7 days. Samples were discarded and reported as negative if CPE were not evident within 14 days. West Nile virus isolates were identified by reverse transcriptase polymerase chain reaction as described by Gibbs et al. (2005).

Spatial Analysis tests

The field of spatial epidemiology includes a wide range of statistical methods (Waller and Gotway 2004), and deciding which test to apply depends on the type of data to be analyzed, the question that needs to be answered, and the validity of the data to meet a particular test' assumptions. Spatial statistical tests can be broadly grouped into three categories: global, local and focused (Ord and Getis 1995; Waller and Gotway 2004). Global tests evaluate spatial patterns and provide a summary of the degree of spatial association among observations over the entire study area. Local tests provide a measure of spatial association between each individual observation (x_i) and its neighbors (x_j) up to a specified distance d. Focused tests are local tests modified to evaluate spatial association only around a known observation (x_i).

The key for most spatial analysis tests relies in the way that spatial structure between observations is incorporated during the inference procedure. The most common way to do this is by generating a spatial weights matrix (W). In this matrix, the values of a cell (w_{ij}) reflect the spatial relationship between observations. w_{ij} can take many forms (Waller and Gotway 2004), but perhaps the two most commonly used forms are based on contiguity or distance thresholds

(i.e., observations within a distance, *d*, or sharing a boundary are considered neighbors and w_{ij} equals 1, otherwise w_{ij} equals 0) or on absolute distances (i.e., the weight entered into cell *ij* is the inverse of distance *d* between observations *i* and *j* raised to the b power, or $w_{ij} = 1/d_{ij}^{\alpha}$, $\alpha \geq 1$). Hence, the values of w_{ij} can be binary (0,1) or continuous.

Below, we describe the statistical tests employed in our article.

Getis and Ord $G_i^*(d)$ Local Statistic (Ord and Getis 1995). In our study we used Getis *Gi*(d)* to identify individual members of clusters of *Cx. quinquefasciatus* abundance and maximum likelihood (ML) WNV infection rates in *Cx. quinquefasciatus*. These local statistics are additive in that the focus is on the sum of the *j* values in the vicinity of point *i*. Hence, we take each trap as a center, one at a time, and search the nearby area for occurrences of more or fewer adult mosquitoes (or ML WNV infection rate in mosquito pools) than expected by random. In this way, specific trap locations are identified as members or non-members of clusters. This statistic is written as

$$G_i^*(d) = \frac{\sum_j^n w_{ij}(d) y_j - W_i^* \bar{y}}{s\{[(nS_{1i}^*) - W_i^{*2}]/(n-1)\}^{1/2}} \quad , \ all\ j$$

where $W_i^* = W_i + w_{ii}$, $S_{1i}^* = \sum_j^n w_{ij}^2$, *all j*

and $\bar{y}$ and *s* are the mean and standard deviation, respectively.

The G_i^* statistic includes the value y_i in its calculations (weighted as 1 in the term w_{ii}) and is asymptotically normally distributed as *d* increases (Ord and Getis 1995). In our study we

used a binary weight w_{ij} based on a distance threshold (d) scheme. The significance of each local Gi^* value at distance d was evaluated using a Monte Carlo randomization procedure. We evaluated the value of Gi^* up to 3 km from each trap. The observed Gi^* was compared with the expected value under 999 Monte Carlo randomizations by using the commercial software ClusterSeer2 (Terraseer, Ann Arbor, MI). Hotspots identified by these statistics (points with Gi^* values significantly higher than the MC generated values) can be interpreted as clusters of high mosquito abundance or ML WNV infection rate.

Bernoulli spatial scan test (Kulldorff 1997). This test evaluates the degree of clustering of events (e.g., WNV positive trap locations) relative to the occurrence of non-events (e.g., WNV negative trap locations). The test is based on a moving, varying diameter, window that evaluates the presence of clusters of events within it. For each window location and size, the number of observed and expected observations inside the window, and a likelihood function (LF) are calculated (the explicit form of the LF for the Bernoulli model is developed in (Kulldorff 1997)). The first step of the method is to find the window which maximizes the LF. After finding the most likely cluster candidates, it is necessary to assess the level of significance of these clusters by using 999 Monte Carlo simulations (Kulldorff 1997). If the considered zone is included within, for instance, the 5% of the replications, the null hypothesis is rejected and a significant cluster is present in that zone. The program SatScan (freely available at http://www.satscan.org/) developed by Martin Kulldorff was used for the Bernoulli Spatial Scan test.

Local Indicators of Spatial Association – LISA (Anselin 1995). The LISA statistic was applied to identify clustering of high human WNV incidence rates and WNV-positive corvid death ratios per census tract within Fulton County. The test statistic is a local decomposition of the classic *Moran's I* test (Anselin 1995), and allows the detection of clusters of similar or dissimilar disease frequencies. Local Moran's I_i is defined as:

$$I_i = \frac{y_i - \bar{y}}{\sum_i^n (y_i - \bar{y})^2 / n} \sum_j^n w_{ij}(y_i - y_j) \quad , \quad i \neq j \ , \ for\ j\ within\ d\ of\ i$$

and the factor of proportionality is:

$$\gamma = \sum_i^n \sum_j^n w_{ij} \sum_i^n (y_i - \bar{y})^2 / n$$

The expected value of I_i is $E(I_i) = -\sum_j^n w_{ij} /(n-1)$. Tests for spatial autocorrelation may be carried out either using the moments of the I_i distribution or by random permutations. Here, we compared the observed I_i with the expected I_i after 999 Monte Carlo permutations. The LISA statistics are particularly useful for identifying spatial clusters. In our study, spatial weights for the LISA test were based on a queen contiguity structure by assuming that areas (e.g., census tracts) that shared a common boundary were direct neighbors (and hence had a w_{ij} value of 1). The program GeoDa (freely available at http://geodacenter.asu.edu/) was used to perform the LISA tests.

Global K-function (Ripley 1976). The global K-function describes the number of pairs of observations between a point, which is the center of a disk and other points that are distance *d* away. For a stationary, isotropic process, $\lambda(d)$ is the expected number of points within distance *d* of an arbitrary point. The estimator of λ is *N/A* where *N* is the number of points in the study area *A*. The estimator of *K(d)* is

$$\hat{K}(d) = \frac{A}{N^2} \sum_{i}^{n} \sum_{j}^{n} \frac{w_{ij}}{e_{ij}} \quad , i \neq j$$

Within the study region of size *A*, for distance *d* we count all of the pairs of points that are $\leq d$ apart. The *W* matrix is made up of 1s for *ij* pairs within *d* of one another and 0 otherwise. As distance increases, the boundary of the region is more likely to be closer to a point *i* than to any *j* point. In that case, an edge correction e_{ij} is included in the formula. A square-root scale makes the function linear and stabilizes the variance, and turns the estimate of *K(d)* as:

$$\hat{L}(d) = \sqrt{\frac{\hat{K}(d)}{\pi}}$$

The expectation of $L(d)$ given the hypothesis of complete spatial randomness (CSR) is *d.* A clustered pattern occurs when $\hat{L}(d)$ is greater than *d,* and lies outside of an envelope containing 999 Monte Carlo permutations of the location of the *N* objects (where each permutation is based on the CSR assumption). We applied the Global K-function to assess whether the location of traps differed from a random distribution with the software ClusterSeer2 (TerraSeer, Ann Arbor, MI).

Kernel density estimation (Silverman 1986). A kernel is a three-dimensional function which weights the events within its sphere of influence according to their distance from the point at

which the intensity is being calculated. Formally, if *s* represents a vector location anywhere in R and $s_1, \ldots, s_n$ are the vector locations of the n observed events (i.e. trap sites), then the intensity $\lambda(s)$, at *s* is estimated as:

$\lambda(s) = \sum_{i=1}^{n} \frac{1}{\tau^2} k\left(\frac{s - s_i}{\tau}\right)$, where $k(\ )$ represents the kernel weighting function (expressed in standardized form). The parameter τ is referred as bandwidth and its value determines the degree of smoothness of the density estimates (high τ values give very smooth surfaces and low τ values give fragmented surfaces). In essence, τ controls for the distance up to which events will have a weight in the estimation of the density at location *s*. In our study we chose a value of λ equal to 1.5 km. Kernel functions were applied to estimate the densities of Cx. quinquefasciatus and of WNV-positive trapping locations using the Spatial Analyst extension of the GIS program ArcGIS 9.2 (ESRI, Redlands, CA).

Supplemental Results

Mosquito collections

In total, 71,824 mosquitoes (totaling 35 species in 10 genera) were collected in 2,041 trap-nights during 2001-2007; 55.1% of all mosquitoes were identified as *Cx. quinquefasciatus*. The odds of WNV infection in the 7,233 species-homogenous mosquito pools tested during 2001-2007 was significantly high for *Cx. quinquefasciatus* collections (OR, 95% CI; 48.1, 6.7-343.6; $P < 0.001$) but not for any of the remaining 34 species (data not shown). Gravid traps collected significantly more *Cx. quinquefasciatus* females (Supplemental Material, Figure 1; Wilcoxon signed-rank test; $z = 7.4$; $P < 0.001$) and detected 11.2 times more WNV infected pools than the paired CDC-light traps. Henceforth, only data for *Cx. quinquefasciatus* collected from gravid traps was used in this study.

A total of 455 trapping locations were surveyed within Fulton County for at least one year during 2001-2007 and were used in our analysis. A global K-function (*L(d)* was applied to the location of traps to determine if their spatial distribution differed from random. The spatial distribution of traps within Fulton County was not random (Figure 2 and Supplemental Material, Figure 2A), but proportional to the population distribution (more traps were placed within the bounds of the city of Atlanta than in the more rural and less populated areas of the county). Clustering of traps occurred at all tested distances, as the observed *Li(d)* was higher than the 999 Monte Carlo randomization bounds (Supplemental Material, Figure 2A). When only traps within Atlanta were analyzed, their distribution did not differ from random at any distance tested, as the

observed *Li(d)* occurred within the bounds of the 999 Monte Carlo Randomizations (Supplemental Material, Figure 2B).

Temporal distribution of ML WNV infection rates in *Cx. quinquefasciatus* (measured as density of infected pools and presence of WNV infected pools) from 2001 to 2006 is shown in Supplemental Material, Figure 3. The year 2007 was excluded because of the low number of mosquito trapping locations monitored. As observed for the overall (2001-2007) pattern (Figure 2), WNV infection in mosquito pools was consistently found in traps located in proximity of CSO-affected streams (Supplemental Material, Figure 3). The proportion of WNV-infected *Cx. quinquefasciatus* pools was significantly higher within 1 km of a CSO stream (125/1215) than of a non-CSO stream (77/1044) ($\chi^2 = 4.9$; P=0.027).

Unsmoothed human incidence

Similar clustering zones to the EB smoothed human WNV incidence rates (Figure 2) but with fewer tracts (2 in the northern and 3 in the southern cluster) were detected when we performed the Moran's I LISA test with the unsmoothed human incidence data (Supplemental Material, Figure 4), validating the findings with the EB smoothed incidence.

Dear bird surveillance

A total of 516 dear bird carcasses were submitted to Fulton County Department of Health and Wellness (FCDHW) during 2001-2007. Most (98.2%) of the WNV-infected birds were

corvids (74.3% blue jays and 23.9% American black crows) (Supplemental Materials, Table 1). Given the high WNV infections detected (327 out of 516 birds tested positive) and that corvid infection is a significant predictor of WNV infection in mosquitoes (Nielsen and Reisen 2007) we only used WNV-positive corvid carcasses in our analysis.

Supplemental References

Anselin L. 1995. Local indicators of spatial association - LISA. Geographical Analysis 27:93-115.

Gibbs SEJ, Mead DG, Moulton JK, Allison AB, Howerth EW, Ellis AE, and Stallknecht DE. 2005. West Nile virus detection in the organs of naturally infected Blue Jays (Cyanocitta cristata). J Wildlife Dis 41: 354-362.

Kulldorff, M. 1997. A spatial scan statistic. Communications in Statistics - Theory and Methods 26:1481-1496.

Nielsen, C. F. and W. K. Reisen. 2007. West Nile virus-infected dead corvids increase the risk of infection in Culex mosquitoes (Diptera:Culicidae) in domestic landscapes. J Med Entomol 44:1067-1073.

Ord, J. K. and A. Getis. 1995. Local spatial autocorrelation statistics. Distributional issues and an application. Geographical Analysis 27:286-306.

Ripley, B. D. 1976. Second order analysis of stationary point processes. J Appl Prob 13:255-266.

Silverman, B. W. 1986. Density estimation for statistics and data analysis. London; New York, Chapman and Hall.

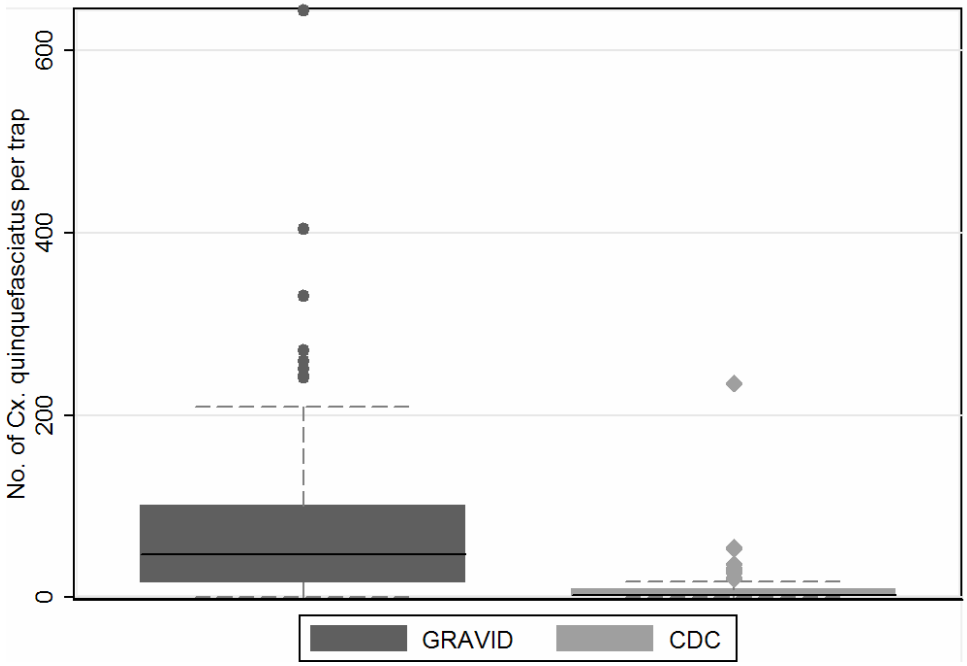
Waller, L. A. and C. A. Gotway. 2004. Applied spatial statistics for public health data. Hoboken, NJ, John Wiley & Sons.

Supplemental Material, Table 1. Number and species name of the 327 WNV-infected bird carcasses identified during 2001-2006 (from a total of 516 birds tested). In the year 2007 no birds were submitted for testing.

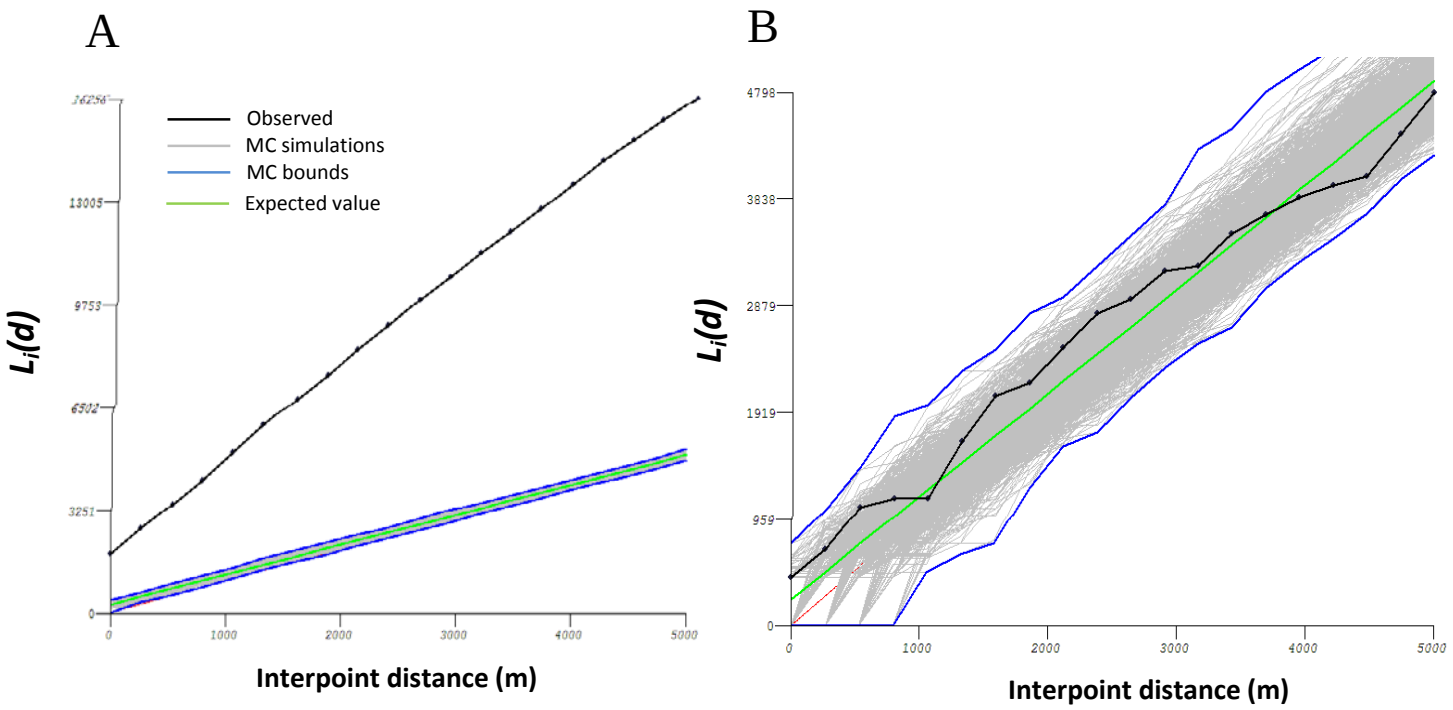
	Reporting year						Grand
Common name	2001	2002	2003	2004	2005	2006	Total
Blue Jay	33	187	14	8	1	0	243
American Crow	16	55	5	2	0	0	78
Fish Crow	0	1	1	0	0	0	2
Cooper's Hawk	0	0	1	0	0	0	1
Eurasian Eagle Owl[1]	0	0	0	0	0	1	1
Red-shouldered Hawk	0	1	0	0	0	0	1
Tufted Titmouse	0	0	0	1	0	0	1
Total	49	244	21	11	1	1	327

[1] Sample provided by Zoo Atlanta.

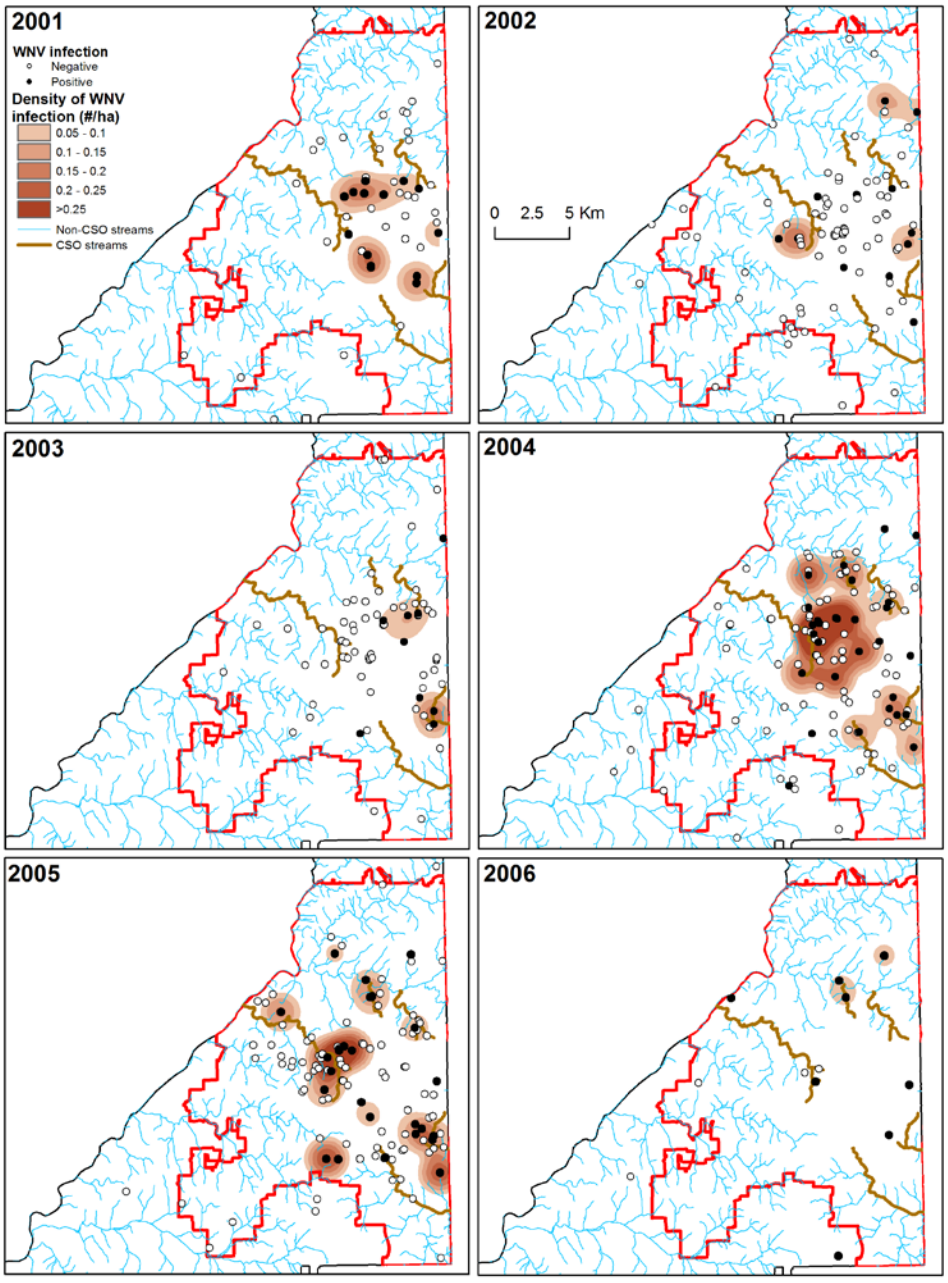
Supplemental Material, Figure 1. Box-plot comparing the median (Q1-Q3, range) number of *Cx. quinquefasciatus* collected in paired Gravid and CDC-light traps from Fulton County (Georgia, USA) from 2001 to 2007.



Supplemental Material, Figure 2. Results from Local-K function ($L_i(d)$) applied to the position of traps within Fulton County (A) and within urban Atlanta (B). Traps are significantly clustered at a given distance (d) when the observed Li lies outside of the confidence bounds for the 999 Monte Carlo Randomizations. Refer to Supplemental Methods and Results for more details.



Supplemental Material, Figure 3. Temporal distribution of mosquito WNV infection per sampling location during 2001-2006 in the city of Atlanta, GA. Infection was represented as presence of WNV-positive pools on a given sampling location (points) and density of infection prevalence (positive pools/tested pools) per hectare (surfaces). Red line indicates the boundary of the city of Atlanta.



Supplemental Material, Figure 4. Spatial clustering (*LISA* map) of the unsmoothed WNV human incidence rates for the city of Atlanta, GA, during 2001-2007.